GTN-Québec
Groupe de travail québécois sur les normes et standards
en TI pour l'apprentissage, l'éducation et la formation

Publication 2010-01

Évaluation de fonctionnalité de traitement des métadonnées par Alfesco en comparaison avec Normetic

---

François Vincent

Mission du GTN-Québec

La mission du Groupe de travail québécois sur les normes et standards pour l'apprentissage, l'éducation et la formation (GTN-Québec) est de fournir une expertise à la communauté éducative en matière de normalisation.

Les membres du GTN-Québec proviennent des trois ordres d'enseignement, des ministères, ainsi que du secteur privé de la formation. En s'appuyant sur les travaux des groupes internationaux d'élaboration des normes, ils soutiennent les acteurs du milieu de l'éducation pour favoriser l'implantation de pratiques communes de description et de production de ressources éducatives interopérables, réutilisables et accessibles à tous.

Ces ressources forment un patrimoine éducatif d'une valeur inestimable pour les communautés éducatives francophones. Assurer son enrichissement et sa pérennité est en conséquence, depuis sa fondation, au cœur des préoccupations du GTN-Québec.

Objectifs du GTN-Québec

1. Dans une perspective d'accompagnement, consulter les acteurs du milieu de l'éducation pour mieux définir comment les approches basées sur les normes et standards peuvent aider à concrétiser la mission éducative de leur organisation ;

2. Connaître des solutions basées sur des normes et standards, s'assurer qu'elles correspondent à la réalité et aux besoins du milieu et proposer, le cas échéant, des adaptations ou des guides d'utilisation de ces normes;

3. Faire connaître et encourager les pratiques normalisées de production et de description de ressources éducatives ;

4. Favoriser le développement d'une masse critique de REA numériques accessibles, pérennes et réutilisables au sein des établissements de chaque ordre d'enseignement ;

5. Maintenir l'expertise et la représentation québécoises en matière de développement de normes internationales et d'autres standards.

Les activités du GTN-Québec sont réalisées avec l'appui financier du ministère de l'Éducation, du Loisir et du Sport du Québec et grâce à la collaboration de ses membres.

www.gtn-quebec.org

ISBN 978-2-924168-08-0 (PDF)

Dépôt légal – Bibliothèque et Archives nationales du Québec, 2012
Dépôt légal – Bibliothèque et Archives Canada, 2012

Sommaire

Le GTN-Québec a envisagé la possibilité de créer des fiches LOM à partir des métadonnées déjà présentes dans les documents. Cette option nous avait été suggérée par l'emploi des métadonnées dans le système de gestion de contenu [Alfresco][1]. Le GTN-Québec a donc commandé cette première étude *Évaluation de fonctionnalités de traitement des métadonnées par Alfresco en comparaison avec Normetic* à WebConforme. Cette étude examine les métadonnées repérées par Alfresco dans divers types de documents, afin de voir lesquelles sont utilisables. Toutefois, une étude plus approfondie nous a montré qu'Alfresco utilisait l'outil [ExifTool][2] de Phil Harvey pour extraire les métadonnées, et n'en exposait qu'une partie: une deuxième étude *Rapport de travail sur ExifTool*, succédant à celle-ci, a donc été réalisée par le secrétariat technologique. On y identifie également la possibilité d'employer le format de métadonnées extensible d'Adobe [XMP][3]. L'approche semblant prometteuse, le secrétariat technologique a également développé un outil permettant de convertir des métadonnées d'un fichier ou d'une URL en une fiche LOM (voir l'outil).

1 http://www.alfresco.com/fr/

2 http://www.sno.phy.queensu.ca/~phil/exiftool/

3 http://www.adobe.com/products/xmp/index.html



GTN-Québec

Évaluation de fonctionnalités de traitement des métadonnées par Alfresco en comparaison avec Normetic

Rapport final

Rédigé par Vincent François, WebConforme

Nº de projet – 0909-03

Montréal, 29 janvier 2010

Table des matières

Révisions du document

Date	Auteur	Changements
11 janvier 2010	Vincent François	Création du rapport préliminaire
15 janvier 2010	Vincent François	Correction
29 janvier 2010	Vincent François	Rapport final

1. Protocole d'analyse

- Compréhension de Normetic 1.2, sa définition, ses exigences, son rapport au LOM *(Learning Object Metadata)*,
- Prise en main d'Alfresco (but, comment, fichiers, métadonnées...)
- Définition des formats à étudier
- Définition du processus de test et tests
- Rapport préliminaire
- Veille sur Alfresco et les extracteurs de métadonnées
- Essai de paramétrisation d'Alfresco et tests additionnels

Formats à étudier

- formats de textes :
  - DOC (Word)
  - ODS (Texte ODF)
  - DOCX (Document Word 2008)
  - TXT
  - HTML
  - XLSX (Classeur Excel 2008)
  - XLS (Classeur Excel)
  - XLS (Classeur Excel 5.0/95)
  - PDF,
  - PDF/A
  - RTF,
  - ODS (classeur ODF),
  - PPT (Présentation Powerpoint),
  - PPTX (Présentation Powerpoint 2008),
  - PPS (Diaporama PPS)
  - PPSX (Diaporama PPS 2008)
- formats d'images :
  - JPG
  - GIF
  - PNG
  - TIFF
  - BMP
  - PSD
  - RAW
- formats audio ou vidéo :
  - url vers une adresse Youtube
  - url vers une adresse Dailymotion
  - WAV
  - MP3
  - MP4
  - MPEG
  - AVI
  - WMV
  - MOV (Quick Time)
  - FLV
  - SWF
  - FLA
  - WMA
  - AIF
  - M4V
- compression :
  - ZIP (arborescence et au moins 2 fichiers)
  - RAR (arborescence et au moins 2 fichiers)
  - TAR
  - GZ
- Modèles de document :
  - DOT (Modèle de document Word)
  - DOTX
  - XLT (Modèle Excel)
  - XLSB (Classeur binaire Excel 2008)
  - XLTX (Modèle Excel 2008)
  - OTT (modèle de texte ODF),
  - OTS (modèle de classeur ODF),
  - POT (Modèle de conception Powerpoint),

  - POTX (Modèle de conception Powerpoint 2008),

Définition du test

Pour chaque format :

- créer un fichier de test,
- le remplir de métadonnées explicites qui puissent être facilement retrouvées et identifiées pour ce qu'elles sont hors du contexte du document (ex. : « nom du document » pour le nom du document, « Auteur principal » pour l'auteur principal...)
- si le format s'y prête, préparer une version française et une version anglaise du même document,
- le sauvegarder dans Basecamp (comme fichier dans la catégorie « Contenu du site » https://webconforme.basecamphq.com/projects/3688339/files),
- le charger dans Alfresco : Espace personnel / Tests Webconforme
- repérer les méta-données qui en ont été extraites par Alfresco,
- ajouter au rapport des formats en tenant compte :
  - des méta-données perdues
  - des méta-données modifiées (troncage, encodage, différence de casse...)
  - des méta-données attendues par Alfresco mais non présentes dans le document.

2. Rapport final

Rapport de formats

Le *rapport des formats* est un tableau, fourni en annexe, présentant les résultats de tous les tests sur tous les formats.

Toutes les métadonnées testées sont présentées, qu'elles fassent partie ou non du profil Normetic et qu'elles soient reprises ou non par Alfresco.

Les formats de fichiers sont regroupés par famille de formats.

Au vu du rapport de format il apparaît les points suivants.

Présence des éléments Normetic 1.2 dans Alfresco

	Requis	Requis Conditionnel	Recommandé	Facultatif	Total
Éléments NE 1.2	19	3	10	26	**58**
Éléments repris par Alfresco	4	1	0	0	**5**
	21%	*33%*	*0%*	*0%*	***9%***

Les éléments Normetic 1.2 sont peu nombreux à être présentés par Alfresco. Seuls les éléments suivants, obligatoires :

- 4.1 Format,
- 1.2 Titre,
- 1.4 Description

sont pris en compte.

Deux autres éléments :

- 1.5 Mot-clé, obligatoire conditionnel,
- 1.3 Langue, obligatoire

sont présents dans les les métadonnées de certains formats de fichier, mais ne sont pas récupérés lors du chargement du fichier dans Alfresco.

Conservation des métadonnées

4.1 Format

Le format est bien récupéré lors du chargement par Alfresco.

Pour les formats texte, le format est bien identifié sauf pour XLSX (Excel 2007), PPS (Powepoint), PPSX (Powerpoint 2007).

Pour les images, seul RAW pose problème.

Pour les fichiers compressés, seul le ZIP passe.

Les audio et vidéo Quicktime AIF, M4V, MOV, MP3, WMA ne sont pas passés.

1.2 Titre

Les divers formats de texte traitent correctement le titre. Le fichier HTML utilise le contenu de la balise *title* du *head*. Seul le format PPS (MS Powerpoint) perd son titre.

Pour les images, deux formats fonctionnent correctement : TIF et JPG. Le format PSD perd son titre.

Aucun format d'audio ou de vidéo ne transmet son titre.

Pour les modèles de documents, seul l'OTS (Open Office Calc) passe son titre.

1.4 Description

Les formats textes transmettent une description depuis des métadonnées différentes :

- *objet* pour les fichiers Open Office Word, Excel,
- *sujet* pour les Microsoft Office Writer, Calc, Powerpoint (sauf pour Powerpoint 2007 qui ne le permet pas), PDF,
- balise *meta description* pour HTML

Les images passent leur description sous la forme d'une métadonnée *objet* pour TIFF et JPG. Le PSD perd sa description en passant sur Alfresco.

Pour les modèles, seul l'OTS (Open Office Calc) passe sa description sous forme de la métadonnée *sujet*.

Suite

Extracteurs de métadonnées

Aujourd'hui, certains formats de fichiers (par exemple, Microsoft Word) supportent directement les modèles de métadonnées simple. Alfresco offre la possibilité d'extraction des métadonnées pour certains de ces types de documents populaires automatiquement.

http://wiki.alfresco.com/wiki/Metadata_Extraction

Ajout de propriétés supplémentaires

D'autres propriétés peuvent être ajoutées au contenu de trois manière :

Une première façon consiste à étendre le modèle de données et définir des propriétés de plus dans un type de contenu.

La seconde consiste à attacher dynamiquement les propriétés et les comportements par *aspect*. En utilisant des aspects, vous pouvez ajouter des propriétés supplémentaires telles que « Effectivity », « Dublin Core Metadata », et « Thumbnailable » au contenu.

http://www.packtpub.com/article/managing-content-in-alfresco

La troisième consiste à utiliser XMP (*eXtensible Metadata Platform*, http://www.adobe.com/products/xmp/)

XMP fournit un format standard pour la création et la gestion des métadonnées (metadata).

XMP est une technologie d'Adobe permettant d'ajouter des données relatives à un fichier directement dans le fichier lui-même, et ce sous forme de méta-données. Les données ainsi incluses deviennent donc inséparables du document auquel elles correspondent, et circulent avec le document lui-même à l'intérieur ou à l'extérieur de l'organisation, au travers des *workflows* et des réseaux. Les données ainsi définies peuvent être lues, modifiées, et enrichies tout au long du processus et du cycle de vie du document, et ce, indépendamment des plates-formes utilisées. (source : http://www.alkero.com/Livre_blanc_XMP.pdf)

Alfresco possède deux extensions XMP qui permettent la manipulation de données XMP en extrayant les métadonnées par un extracteur XMP et en ajoutant des données directement dans le fichier *(embeder).*
http://forge.alfresco.com/projects/xmp
http://forge.alfresco.com/projects/embedders

1. ANNEXE

Rapport des formats *« rapport des formats.xls »*

Publications du GTN-Québec

2012-03	*Soutien au développement de ressources numériques pour l'enseignement et l'apprentissage dans les universités québécoises – Rapport complet.* Rédigé par Line Cormier, Maureen Clapperton, Nicolas Gagnon, Michel Gendron, Robert Gérin-Lajoie et Jean Marcoux, 71 p.
2012-02	*Soutien au développement de ressources numériques pour l'enseignement et l'apprentissage dans les universités québécoises – Les faits saillants.* Rédigé par Line Cormier, Maureen Clapperton, Nicolas Gagnon, Michel Gendron, Robert Gérin-Lajoie et Jean Marcoux, 10 p.
2012-01	*Manuels de cours numériques – droit d'auteur et gestion, inventaire des solutions disponibles version 1.1.* Rédigé par Réjean Payette, 38 p.
2011-06	*Les tableaux numériques interactifs : considérations d'interopérabilité.* Rédigé par Marc-Antoine Parent, 28 p.
2011-05	*Fédération d'identité pour les organismes de l'éducation.* Rédigé par André Breton, 50 p.
2011-04	*Compte-rendu de participation, 26ème colloque annuel CSUN 2011.* Rédigé par Denis Boudreau, 14 p.
2011-03	*Les environnements d'apprentissage sont-ils en mutation ou en gestation?* Rédigé par Pierre-Julien Guay, Marcel Borduas, Yves Otis, Robet Paré et Sacha Leprêtre, 21 p.
2011-02	*Profil d'application québécois de métadonnées pour les opportunités d'étude, d'apprentissage et de formation (v.0.7.5)* Rédigé par Gilles Gauthier, 93 p.
2011-01	*Profil d'application Normetic 2.0 (v0.7.5)* Rédigé par Gilles Gauthier, 41 p.
2010-01	*Évaluation de fonctionnalités de traitement des métadonnées par Alfesco en comparaison avec Normetic.* Rédigé par François Vincent, 9 p.
2009-06	*Portrait des pratiques de sélection, de catalogage et de partage des documents numériques dans les bibliothèques.* Rédigé par Marie-Chantal Dufour, 48 p.
2009-05	*Accès aux contenus de formation en ligne : difficultés des apprenants handicapés et solutions pour assurer l'accessibilité des contenus.* Rédigé par Denis Boudreau, 21 p.
2009-04	*Développement MLO: Metadata for learning opportunities.* Rédigé par Olivier Gerbé et Thi-Lan-Anh Dinh, 32 p.
2009-03	*Concept and Prototype of an Aggregator Portal for Learning Opportunities Based on the MLO-AD Standard.* Rédigé par Katharina Bauer-Öppinger, 89 p.

(autres publications à la quatrième de couverture)

Publications du GTN-Québec (suite)

2009-02	*Identification des caractéristiques des modèles de diffusion de contenus numériques : recension des dépôts numériques existants – Partie 2.* Rédigé par Gabriel Dumouchel et Thierry Karsenti, 99 p.
2009-01	*Identification des caractéristiques des modèles de diffusion de contenus numériques : revue de littérature – Partie 1.* Rédigé par Gabriel Dumouchel et Thierry Karsenti, 54 p.
2008-05	*Ressources d'apprentissage et normes : la situation au Québec.* Rédigé par Christian Lafrance, 102 p.
2008-04	*Guide d'élaboration de fiches descriptives de ressources d'enseignement et d'apprentissage selon Normetic v1.2, profil d'application québécois du standard Learning Object Metadata (LOM).* Rédigé par Gérald Roberge, 57 p.
2008-03	*Profil d'application Normetic 1.2*. Rédigé par Gérald Roberge, 170 p.
2008-02	*Tableau du code XML à produire pour le vocabulaire de l'élément 5.2 de Normetic 1.2*. Rédigé par Gérald Roberge
2008-01	*Tableau du code XML à produire pour le vocabulaire de l'élément 5.6 de Normetic 1.2.* . Rédigé par Gérald Roberge
2007-01	*Portrait général des stratégies d'assurance qualité des ressources d'enseignement et d'apprentissage (REA) : à l'attention des gestionnaires.* Rédigé par Karin Lundgre-Cayrol, Suzanne Lapointe et Ileana De la Teja, 25 p.
2006-03	*Les normes, comment?* Rédigé par Gérald Roberge, 4 p.
2006-02	*Les normes, pourquoi?* Rédigé par Gérald Roberge, 4p.
2006-01	*Guide pour la sélection de REA.* Rédigé par Gérald Roberge, 10 p.
2005-01	*Le profil d'application Normetic, version 1.1.* Rédigé par Robert Thivierge, 8 p.
2003-01	*La description normalisée des ressources : vers un patrimoine éducatif – Normetic, version 1.0.* Sous la supervision de la CREPUQ et Novasys inc., 139 p.
2002-01	*Les normes et standards de la formation en ligne – État des lieux et enjeux.* Rédigé par Rachel Chouinard. Sous la supervision de la CREPUQ et du sous-comité SCTIC, 39 p.

Pour télécharger ces publications ou pour la liste complète des publications du GTN-Québec, voir le site Web **www.gtn-quebec.org/publications**